Research Article

Huanhuan Zhang and Yang Yu*

Research on pattern recognition of tourism consumer behavior based on fuzzy clustering analysis

Abstract: Understanding customer behavior has become critical for marketers and tourist service providers since global travel has rapidly expanded. Recognizing consumer behavior patterns allows organizations to modify their strategy for better service delivery. However, the conventional approaches frequently fail to adequately capture the variety and complexity of tourism consumer behavior because of the large and diverse data available. This research seeks to investigate the use of fuzzy clustering analysis to better understand tourism consumer behavior patterns. The method combines fuzzy clustering algorithms with customer behavior data such as demographics, travel preferences, and purchasing patterns. The investigation reveals separate groups of consumers, providing insights into how various factors influence tourist purchasing decisions. The data were gathered using questionnaires, online booking platforms, and travel websites, where customers provided information about their previous travel experiences and preferences. Data preparation was used to normalize the data for analysis. Principal component analysis was employed to decrease dimensionality. The Sea Turtle Foraging Optimized Fuzzy C-Means clustering (STFO-FCMC) is presented as an extension of normal FCMC that incorporates an optimization procedure based on sea turtle foraging habits. This optimization enhances the accuracy and efficiency of cluster center selection and membership values, making STFO-FCMC especially well-suited for dealing with the complexity and unpredictability of tourism behavior data. The findings show multiple consumer behavior patterns, including diverse preferences for various types of tourist products and services, which are split by age, income, and travel objectives.

The STFO-FCMC method is assessed using metrics, including accuracy of 97.84%, precision, recall, and F1-score. These data assist service providers create individualized services and marketing strategies that improve consumer satisfaction and business performance. Overall, fuzzy clustering analysis, particularly with the STFO-FCMC approach, is a successful tool for detecting tourist consumer behavior, with substantial promise for improving tourism product and service targeting.

1 Introduction

Tourism has evolved into one of the most important worldwide businesses, considerably contributing to countries' economic and social growth [1]. As the tourist sector grows, understanding customer behavior has become increasingly important. Consumer behavior includes obtaining and organizing information to make purchase decisions, as well as analyzing products and services. The technique involves looking, acquiring, utilizing, evaluating, and discarding things and services. Tourist purchases have two distinct characteristics: they are not a measurable investment, and they are generally the result of long-term savings [2]. Tourism consumer behavior is an assessment of how individuals, groups, or associations choose, buy, use, and organize tourism-related products or services [3]. A consumer's individualized experience that corresponds to their particular preferences is more favorable than a comprehensive expression of their needs. Understanding these behaviors is critical to tourism stakeholders, including travel agents, hotel operators, tour operators, and destination marketers [4]. Through monitoring consumer behavior, these stakeholders devise more successful marketing tactics, tailor their services to specific market segments, and increase customer satisfaction [5]. Cluster analysis

* **Corresponding author: Yang Yu,** School of Tourism, Xinyang Agriculture and Forestry University, Xinyang, 464000, China, e-mail: zhanghuanhuan2005@163.com
Huanhuan Zhang: School of Tourism, Xinyang Agriculture and Forestry University, Xinyang, 464000, China; School of Education, Huazhong University of Science and Technology, Wuhan, 430074, China

identifies homogenous clusters for describing and grouping multidimensional data. By aligning the units within each group as much as possible for comparability against one another, while maximizing the individuality of the groups themselves, effective market segmentation, in a rudimentary sense, at least, began back in 1950 and has covered the way towards several other clustering methodologies [6]. The improvement of digital platforms, such as social media, online reviews, and travel websites, has given much data for research purposes. Hence, each cluster analysis can provide a different image of a whole by testing various clustering algorithms via a multivariate descriptive instrument [7]. However, these data are often complex, unstructured, and high-dimensional, posing challenges for conventional data analysis methods. The objective of this research is to explore the application of fuzzy clustering analysis in recognizing patterns of tourism consumer behavior, identifying distinct segments, and providing insights to enhance marketing strategies and service offerings within the tourism industry.

1.1 Key contribution

- The objective is to explore the application of fuzzy clustering analysis in recognizing patterns of tourism consumer behavior, identifying distinct segments, and providing insights to enhance marketing strategies and service offerings within the tourism industry.
- Initially, data is collected from kaggle. The data is preprocessed, including a robust scalar, and then features are extracted using principal component analysis (PCA) to reduce dimensionality.
- Sea Turtle Foraging Optimized Fuzzy C-Means clustering (STFO-FCMC) is proposed to enhance clustering efficiency by optimizing cluster centers and membership values.
- This approach effectively identifies distinct consumer behavior patterns, offering valuable insights for personalized marketing strategies and improved service delivery.

1.2 Organization of this work

This research is structured as follows: Section 2 explains related work, Section 3 explains the method workflow, Section 4 evaluates the result, Section 5 describes the discussion, and Section 6 concludes the research.

2 Related work

This section covers tourist customer behavior utilizing artificial intelligence (AI) and the Internet of Things (IoT). Outlining the pros and cons of the techniques used. It also expands on the theoretical and practical framework established by previous research in this sector.

Ma [8] enhanced smart tourism service systems by integrating IoT and machine learning (ML) approaches. The method involved evaluating the system performance through a simulated dataset to assess congestion detection in tourist areas. The finding displayed that the system outperformed a support vector machine (SVM) based approach in prediction accuracy. The shortcoming was the use of a simulated dataset, which did not adequately replicate real-world settings. Saragih et al. [9] explained the role of service robots in tourism, employing a text-mining method based on ML algorithms. Patterns and motifs were investigated using latent Dirichlet allocation models. Eight key themes were identified, emphasizing four potential research fields. Limitations include possible bias in topic modeling and a restricted temporal span. Liu [10] developed a neural network (NN) based accuracy-advertising model to enhance user churn prediction and user value development. The model integrated a data mining approach and conducted empirical tests on product and market strategies within big data platforms. Results suggest that precision marketing must be more detailed and align with user-sensitive information and market conditions. The limitation was real-world implementation validation and does not account for rapidly changing consumer behaviors. Table 1 shows the overview of methods, objectives, results, and limitations in the previous literature review.

Tourism data comes from a variety of structured and unstructured sources, such as online booking records, travel itineraries, GPS trajectories, social media check-ins, sentiment-laden reviews, and environmental sensors. These many data sources are critical to analyzing visitor behavior, preferences, and satisfaction. Several researchers [12–14] have used sentiment-rich text, while others [16,17] used clustering and regression algorithms on spatiotemporal datasets. However, the integration of diverse tourist data, particularly in real-time, remains underexplored, limiting the depth of behavioral insights and the responsiveness of tourism management systems. The approach is trained and assessed using real-world tourist data in a variety of settings, in contrast to previous work that either uses simulated data [8], and region-specific datasets [17], or has poor real-world validation [10,15]. It further improves prediction accuracy and interpretability. Consequently, it not only enhances forecasting but also

Table 1: Overview of literature review

| References	Year	Methods	Objective	Result	Limitations
[11]	2024	ML approach to analyze search behaviors and forecast tourism demand	Examined search data influence on model optimization and accuracy	ML model outperformed other models. More search data does not always enhance accuracy	Findings are specific to Baidu data and require broader validation
[12]	2024	Aspect-based sentiment analysis and deep learning (DL) prediction model	Develop a personalized restaurant recommender system capturing granular preferences	Superior prediction performance over five existing models	Potential evaluation biases and dataset constraints
[13]	2024	DL-based approach, culminating in the development of a DL model for sentiment classification in smart tourism	To classify textual data in smart tourism by conducting sentiment analysis and categorizing reviews as positive or negative	DL model outperforms across multiple evaluation metrics, demonstrating superior accuracy, and $F1$-score, which highlighted its effectiveness in sentiment classification	The model performance was influenced by biased review data, domain-specific constraints, and challenges in handling multilingual or nuanced sentiments
[14]	2023	Models are applied to address data imbalance	To evaluate ML algorithms for sentiment classification in Culture and Heritage Tourism content	Superior ML models enhance sentiment classification in Culture and Heritage Tourism	YouTube comments on content, potentially restricting generalizability across diverse tourism contexts
[15]	2024	ML approaches data analysis applications in the generosity industry, examining their evolution, capabilities, and integration with social media analytics	To explore the transformative latent of ML-driven data analytics in enhancing executive, customer personalization, and effective competence in the hospitality industry	ML-driven analytics significantly enhance demand forecasting, personalized marketing, and predictive maintenance	Limited to existing literature and lacks empirical validation through real-world case studies or experimental data
[16]	2025	Fuzzy clustering and additive logistic regression analysis were applied to a large cross-sectional dataset to identify tourist types and analyze temporal changes in travel behavior	To understand evolving travel behavior patterns and inform tourism strategies	Identification of five tourist types with varying behaviors over age, time, and generations	Potential biases in the dataset limit broader applicability
[17]	2025	Utilize information from the 2019 Puget Sound Regional Household Travel Survey, applying similarity propagation and k-means clustering	Investigate periodical shopping and activity trip patterns of households	Recognized six clusters with diverse trip patterns and sociodemographic differences	Focused on specific regional data, limiting broader generalization of results
[18]	2025	Fuzzy logic system integrating Technique for Order Preference by Similarity to Ideal Solution for customized travel proposals based on concurrent user data	Enhance consumer knowledge in smart tourism via fuzzy logic-based personalization	The system improves user satisfaction through precise, adaptable recommendations	Further evaluation is needed in diverse real-world tourism platforms and scenarios
[19]	2025	A mixed-systems model with regression analysis and fuzzy set assumption was used	The research explored traveler favorites for AI versus human mediators in airline customer service	Findings revealed passengers prefer AI for simple assignments and human mediators for multifaceted issues	Limitations include sample size and context specificity
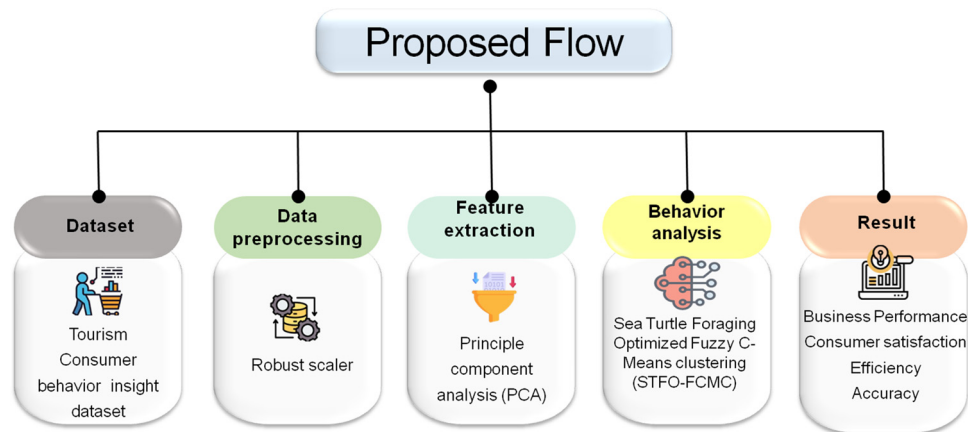
Figure 1: Methodology workflow.

provides context-sensitive suggestions for travel service providers.

3 Methodology

The section initially describes the data collected from the Tourism Consumer Behavior Insights Dataset. After that, the pre-processing technique includes a robust scalar, followed by feature extraction using PCA for dimensionality reduction. STFO-FCMC is employed to enhance clustering efficiency by optimizing cluster centers and membership values. Figure 1 demonstrates the methodology workflow.

3.1 Data collection

The data is gathered from the Tourism Consumer Behavior Insights Dataset in Kaggle [20]. This dataset offers a thorough understanding of customer behavior in the travel industry by integrating demographic, behavioral, and booking-related data. It documents the planning and execution of travel activities by people with varying tastes and budgets. Age, gender, income, kind of travel, and length are among the characteristics included in the data. It also takes into account reviews and reservation methods to replicate actual customer interactions in the travel industry.

3.2 Data pre-processing using robust scalar

Data normalization methods, such as robust scalars, are used to scale features and increase the model's resistance

to outliers. The median and scaling according to the inter-quartile range (IQR) method is used to scale features in a way that is less susceptible to outliers than traditional approaches, such as z-score normalization or min–max scaling. IQR is the range between the first quartile (25th quantile) and third quartile (75th quantile) as represented in the following equation:

$$\mathrm{RS}(Y_j) = \frac{(Y_j - \mathrm{median}\,(y))}{\mathrm{IQR}_{1,3}\,(Y)}, \tag{1}$$

where Y denotes the data point and positive values are indicated using $\mathrm{RS}(Y_j)$. IQR is the specific difference between the data's first and third-quartile values. $Y_j - \mathrm{median}\,(y)$ is the median of data $\mathrm{IQR}_{1,3}(Y)$ metric for a central propensity that is less influenced by random variations.

3.3 Feature extraction using PCA

After pre-processing the data, PCA is employed to reduce dimensionality and highlight the most significant features. PCA transforms a set of parallel features into a rest of unrelated principal compounds, capturing the directions of maximum variance in the data. The data retains its key features, making it easier to analyze patterns in customer preference, travel destinations, and behaviors. PCA is a statistical technique that uses diagonal transformations to convert potential association values into separate variables. It focuses on the modified variables and the reduced set of similar characteristics in high-dimensional data, which is a significant aspect. It reduces the total translation error; this transformation is carried out using PCA by identifying the ∅ feature of reflecting and

vectorizing n-dimensional information into the package. A significant amount of the information present in the initial high-dimensional collection is retained when it is stretched into a lower-dimensional subspace. Let us assume that v_S is a random dimensional N with a mean of $(\varnothing)$ input data recording as

$$\varnothing = \frac{1}{N}\sum_{s=1}^{N} v_s, \tag{2}$$

where v_s is the sth data vector in the dataset, and each v_s is a vector of features. The definition covariance matrix is represented in equation (3), N is the number of sample data, and E is determined by finding eigenvalue and eigenvector $(w_s - \mu)(w_s - \mu)^S$. The data factors level highlights the most influential factors in content consumption

$$E = \frac{1}{N}\sum_{s=1}^{N} (w_s - \mu)(w_s - \mu)^S, \tag{3}$$

where w_s is the sth data sample, μ is the mean vector, S represents the transposition operator, and E is the covariance matrix. PCA solves the variance matrix DDu_s an eigenvalue of λ_s concern using equation (4), a threshold value that ensures the explained variance is sufficiently high

$$DDu_s = \lambda_s u_s, \tag{4}$$

where D is the covariance matrix and λ_s is the eigenvalue. The equivalent eigenvector is λ_s and the eigenvalue is m in equation (5), it assists the dimensionality reduction, with PCA choosing the top m components that most variance in the data

$$\frac{\sum_{s=1}^{n}\lambda_s}{\sum_{s=1}^{m}\lambda_s} \geq u\frac{\sum_{s=1}^{n}\lambda_s}{m} \geq u, \tag{5}$$

where m is the retained components, n is the total integer of constituents, and u is the variance threshold. Customer preferences, travel locations, and habits, with a focus on key aspects such as budget, lodging, and destination type. This dimensionality reduction increased clustering accuracy and model efficiency when recognizing patterns.

There are 23 characteristics and 1,200 samples in the original dataset. These characteristics are chosen since they are pertinent to identifying trends in traveler behavior. The dataset's dimensionality is decreased from 23 to 8 major components using PCA. These eight elements reduced noise and duplication while maintaining the bulk of information, accounting for 91.3% of the overall variance. The STFO-FCMC algorithm is fed the PCA-transformed data that is produced. Figure 2 displays the PCA visualization of three tourist customer clusters.

3.4 Tourism consumer behavior analysis using STFO-FCMC

STFO-FCMC improves cluster centers and membership values through efficient consumer behavior analysis. FCMC effectively handles the uncertainty and vagueness in consumer behavior data, such as mixed preferences for travel destinations or accommodation types. STFO in tourism consumer behavior analysis is drawn as a metaphor for understanding customer decision-making patterns.

3.4.1 FCMC

The dimension reduction data is clustered using FCMC. The unsupervised soft clustering method FCMC enables each data point to have varied degrees of membership in several groups by dividing a dataset into a predetermined number of clusters. FCMC employs a membership function, which usually ranges between 0 and 1, to indicate the degree to which a data point belongs to each cluster rather than allocating each data point to a single cluster. By decreasing an objective function that quantifies the distance between data points and cluster centers, weighted by the degree of membership, the method iteratively updates cluster centers and membership values. FCMC is employed due to its ability to capture the uncertainty and overlaps commonly observed in consumer behavior data. Unlike traditional clustering methods like K-means, which assign each data point to only one cluster, FCMC allows partial membership across multiple clusters. This is particularly important for analyzing travel preferences, where
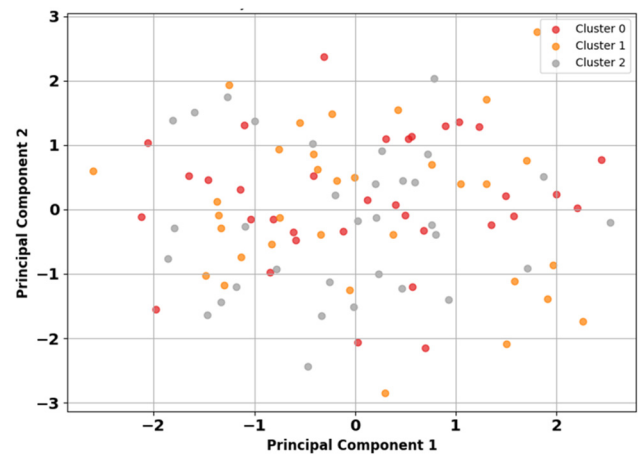
Figure 2: Variance preserved by each component.

tourists exhibit characteristics of more than one segment. FCMC improves classification granularity, enhances interpretability, and enables more nuanced consumer profiling compared to hard clustering algorithms. Fuzzy clustering, such as the fuzzy c-means method, has been utilized to investigate tourist satisfaction by categorizing replies in such a way that overlaps are allowed, reflecting the inherent ambiguity of human attitudes. FCMC is especially well-suited for tourism research as it permits each visitor to be a member of several segments to differing degrees. Travelers display characteristics of both adventure seekers and frugal tourists at the same time. Because of this overlap, which conventional hard clustering techniques are unable to capture, FCMC is useful for simulating intricate, real-world customer behavior. Equation (6) represents the membership matrix $\sum_{i=1}^{M}\sum_{j=1}^{d}v_{ji}^{n}c_{ji}^{2}$, which corresponds to class I_n for each element certain degree. The FCMC method continuously updates the class center and membership matrix on each iteration until the criteria function reaches its minimum

$$I_n(V, U) = \sum_{i=1}^{M} \sum_{j=1}^{d} v_{ji}^{n} c_{ji}^{2}, \tag{6}$$

where $I_n(V, U)$ is the objective function, M is the sum of data points, d is the sum of clusters, v_{ji}^{n} is the membership degree, n is the fuzzification coefficient, and c_{ji}^{2} is the distance between data points. The sum of the distance rectangles among the members to the cluster center is supplied using equation (7). The weight matrix is v_{ji} then Euclidean distance $\left(\frac{c_{ji}}{c_{lj}}\right)$. FCMC uses image iterative object function modification to identify set partitions. The object function is inferred by the sum of the distance squares among data points and the cluster center

$$v_{ji} = \sum_{l=1}^{d} \left(\frac{c_{ji}}{c_{lj}}\right)^{\frac{2}{1-n}}. \tag{7}$$

FCMC optimizes the object function using the initial value, which is highly impacted by the convergence speed, especially in large clusters. Cluster centers identify the best cluster centers (centroid) by iterating over the fuzzy membership matrix and updating the centers depending on the weighted distances between members. Equation (8) updates the partisanship rates v_{ji} of the ith data point to the jth cluster based on the fuzzy intensity

$$v_{ji} = \frac{\sum_{i=1}^{M} V_{ji}^{n} W_j}{\sum_{i=1}^{M} V_{ji}^{n}}, \tag{8}$$

where W_j is the weight value for cluster and n is the fuzzification parameter. It helps to adjust membership degrees

by considering cluster-specific intensities, especially relevant in weighted clustering applications. One cluster represents adventure-seeking travelers, while another corresponds to tourists looking for luxury and leisure. The similarity function, $g(w) = O_0$ a Euclidean matrix, shows the similarity between pixels in FCMC using

$$g(w) = O_0 + \left[1 - \exp\left(-\frac{w}{b}\right)\right], \tag{9}$$

where O_0 is the baseline similarity offset, w is the distance among data points, and b is the decay parameter controlling the steepness of the similarity curve. This function maps spatial proximity to similarity, ensuring that nearby pixels are more likely to belong to the same fuzzy cluster. Once the clusters are created, tourism businesses analyze their group's behaviors and characteristics. For instance, one cluster represents high-spending tourists who enjoy luxury resorts, and another represents budget travelers looking for adventure trips. The membership degrees facilitate the acquisition of qualitative insights into consumer preferences. For $0 \leq w \leq K$, the membership function is $K \leq w \leq d$, indicating the quadratic and w approach up to a minimum value at K.

The variable $d \leq w \leq G$ defines the degree to which a particular element $v(w)$ belongs to the fuzzy set. For example, in tourism consumer behavior modeling, how strongly a particular customer fits into various categories of consumer inclinations is established on a range of behaviors. The membership function, which establishes the extent to which an element w belongs to a fuzzy cluster, is represented by the symbol $v(w)$ in equation (10). A membership value in the interval [0, 1], where 0 indicates no partisanship and 1 describes total partisanship, is mapped to the input variable w. The function is piecewise constructed to capture the ambiguity of consumer cluster assignments by reflecting distinct membership levels across various ranges of w

$$v(w) = \begin{cases} 2\left(\dfrac{w}{d}\right)^2 & 0 \leq w \leq K \\ 1 - 2\left(\dfrac{d-w}{d^2}\right) & K \leq w \leq d \\ 1 - 2\left(\dfrac{w-d}{1-d}\right)^2 & d \leq w \leq G \\ 2\left(\dfrac{1-W}{1-d}\right)^2 & G \leq w \leq 1. \end{cases} \tag{10}$$

This piecewise membership function modifies the degree to which a data point belongs to various clusters according to the value of w. While the breadth of the transition zones between full and no membership is controlled

by the parameters d and G, the parameter K establishes the limit for partial membership. Equation (11) represents the intention function used to assess the inclusive performance of the fuzzy clustering model. $(W_j - W_{j-1})$ denotes the distance metric for the clusters and $v(U)c \in [0, 1]$ data points. $C = 1$ represents the penalty factor that controls the trade-off between differences between fitting the data points and cluster centers in FCMC

$$C = 1 - \sum_{i=1}^{3}(W_j - W_{j-1})v(U)c \in [0, 1]. \qquad (11)$$

The total fuzziness of membership and the distances among the data points and their respective centroid are minimized. Equation (12) denotes the evaluation function to measure the quality of the fuzzy clustering solution. The cluster center is $I(V, U)$ and the fuzzy membership of the data point is $C^2(w_j, u_j)$ raised to the power of n, which is a common way to give more or less weight to memberships based on their strength

$$I(V, U) = \sum_{i=1}^{M}\sum_{j=1}^{d} v_{ji}^n C^2(w_j, u_j). \qquad (12)$$

Equations (13) and (14) determine fuzzy membership computation, which determines how much an observation belongs to a cluster

$$v_{ji} = \sum_{l=1}^{d}\left(\frac{C_{ji}}{C_{lj}}\right)^{\frac{2}{1-n}}, \qquad (13)$$

$$u_j = \sum_{i=1}^{M} v_{ji}^n w_i / \sum_{i=1}^{M} v_{ji}^n. \qquad (14)$$

The fuzzy clustering method assigns membership values to numerous groups depending on distance. To increase the accuracy of grouping overlapping data, such as in tourist consumer behavior research, the membership function and cluster centers could be optimized.

3.4.2 STFO

The STFO algorithm is used to enhance the clustering performance of the FCMC. The STFO is used to address the limitations of FCMC, which often suffers from sensitivity to initial cluster centers and local minimum entrapment. STFO mimics the intelligent foraging behavior of sea turtles, which balance exploration and exploitation based on environmental cues. In this context, the STFO algorithm optimal decision-making in tourist consumer behavior by balancing exploration and exploitation while reaping individualized suggestions from consumer preferences and

environmental conditions. Sea turtles search for optimized strategies according to environmental signals, while tourists display selective behavior influenced by destination attraction, money, and personal choice. Following these trends allows tourist providers to adjust their offers to individual needs and increase engagement and satisfaction. This technique also emphasizes the need for sustainability in tourism, similar to the ecological factors in sea turtle behavior. Specify a starting population of O_j turtles and create each turtle's initial location at random. A population of virtual sea turtles is formed, with each representing a potential solution or choice in the tourism environment, such as a tourist location or package. Generate random beginning velocities for sea turtles' velocity of $[o_1^j, \; o_2^j, \; ..., \; o_C^j]$ as shown in equations (15) and (16). To determine the fitness value of high-density data regions, evaluate its beginning location using the objective function is $\alpha(\text{WVA} - \text{WKA})$. Evaluate each turtle's location using an objective function to determine their fitness value max and min function $U_{\min} = -U_{\max}$ is represented in equation (17). Record the turtle with the greatest fitness value $[L_1^i, \; L_2^i, \; ..., \; L_C^i]$ using equation (18).

$$O_j(0) = [o_1^j, \; o_2^j, \; ..., \; o_C^j], \qquad (15)$$

$$U_{\max} = \alpha(\text{WVA} - \text{WKA}), \qquad (16)$$

$$U_{\min} = -U_{\max}, \qquad (17)$$

$$L_i(0) = [L_1^i, \; L_2^i, \; ..., \; L_C^i]. \qquad (18)$$

The sea turtle has a higher fitness value than the data source; consequently, the data source's contribution is considered zero. If the turtle's fitness value is less than the data source's, the contribution of the supply of data is determined as follows index J that maximizes the value of $(e_{o_{j(s)}})$, indicating the optimal choice based on some evaluation metric e using

$$J = \frac{\arg\max(e_{o_{j(s)}})}{j}. \qquad (19)$$

Choose the turtle's finest data source. The data source with the highest value among all the others is the best. Equation (20) updates the utility $U_j(s - 1)$ for an object system at step s adjusting it based on the change in the evaluation metric $e_{o_{j(s)}} - e_{o_{j(s-1)}}$ between consecutive steps using

$$U_j(s) = U_j(s - 1) + \left(\frac{e_{o_{j(s)}} - e_{o_{j(s-1)}}}{e_{o_{j(s-1)}}}\right)(e_{j(s)} - e_{j(s-1)}). \quad (20)$$

By optimizing utility and effectively finding the best locations or packages, the STFO algorithm improves consumer behavior analysis in the tourist industry and raises customer happiness and engagement.

STFO-FCMC overcomes the challenges by using the global search capability of STFO to generate high-quality initial solutions, thereby improving convergence speed and solution quality. This method also improves clustering accuracy and increases robustness in the presence of noisy, ambiguous data like diverse tourist preferences. Thus, the STFO-FCMC method is selected for its capability to effectively discover the solution gap and provide stable clustering results aligned with real-world tourist segmentation needs. This hybridization leads to more precise insights, enabling tourism providers to better tailor offerings and enhance consumer satisfaction. Pseudocode 1 shows the STFO-FCMC method.

Pseudocode 1: Sea Turtle Foraging Optimized Fuzzy *C*-Means clustering (STFO-FCMC)

Initialize_parameters (*c,n,max_iterations,tolerance*)
Initialize_membership_matrix(U,c,num_data_points)
Initialize_sea_turtle_positions(positions,velocities)
For iteration in range (max_iterations):
Update Cluster Centers (FCMC Core)
For j in range(c):
numerator = np.sum((U[:,j]**n)[:,np.newaxis]
*data,axis=0)
denominator = np.sum(U[:,j] ** n)
cluster_centers[j] = numerator/denominator
Update Membership Matrix
For i in range(num_data_points):
For j in range(c):
distance = np.linalg.norm(data[i] - cluster_centers[j])
U[i,j] = 1 / np.sum([
(distance/np.linalg.norm(data[i]-cluster_centers[k])) ** (2/
(n − 1))
for k in range(c)
])
Sea Turtle Foraging Optimization
For j in range(c):
fitness_current = objective_function(U,cluster_centers)
if fitness_current > best_fitness[j]:
best_fitness[j] = fitness_current
best_positions[j] = cluster_centers[j]
rand_vec = np.random.rand(*cluster_centers[j].shape)
velocity[j] = alpha * (best_positions[j] - cluster_centers[j])
+ beta * rand_vec
cluster_centers[j] = cluster_centers[j] + velocity[j]
If np.linalg.norm(U - previous_U) < tolerance:
break
previous_U = U.copy()
Return cluster_centers,U

The approach effectively enhances tourism consumer behavior analysis, offering improved clustering efficiency and insightful segmentation for targeted marketing strategies. Its scalability ensures that it can accommodate data growth as more variables and observations are included.

4 Result

The goal is to explore the application of fuzzy clustering analysis in recognizing patterns of tourism consumer behavior, identifying distinct segments, and providing insights to enhance marketing strategies and service offerings within the tourism industry. This research uses metrics such as accuracy and loss, performance metrics, cluster analysis on tourism consumer behavior, consumer satisfaction, and efficiency in consumer behavior analysis. Table 2 displays the experimental setup.

4.1 Accuracy and loss

Accuracy is the number of correct predictions made by a classical model to the total number of predictions, whereas loss is the difference between expected and actual values, which measures how well the model performs throughout training. It enables one to evaluate the quality of clustering using external criteria since accuracy gauges how well the projected clusters match the real labels following optimum mapping. In this research, accuracy is a post-clustering validation metric rather than a sign of supervised learning. The accuracy and loss characteristics of the training for the STFO-FCMC technique are shown in Figure 3. The loss curve shows convergence behavior, where lower loss indicates improved model stability and optimization. The accuracy curve provides insights into how well the discovered clusters align with actual tourist behavior patterns, enabling better-informed marketing strategies and personalized service offerings in the tourism sector.

4.2 Performance metrics

The performance metric allows for the evaluation of how well the model solves a specific problem. It determines the predictive recall, accuracy, precision, and *F*1-score of the model in recognition patterns, balancing the correct identification of patterns with reducing the number of false
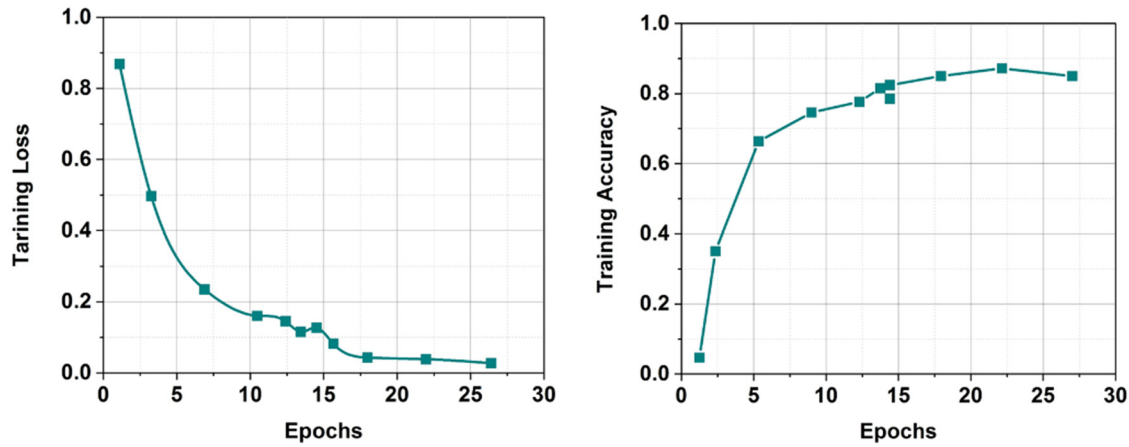
Figure 3: Result of loss and accuracy.

positives and false negatives within the tourism behavior prediction. These metrics give insight into the model's capacity to effectively discover patterns while minimizing false positives and false negatives. The $F1$ score is especially important since it combines accuracy and recall into a single metric, offering a comprehensive perspective of model performance in cases when class distribution is skewed. The proposed method is compared with traditional methods like light gradient boosting (LGB) [21] and stacking [21]. Figure 4 displays the result of the performance metric. Table 3 shows the comparative analysis of the proposed and existing methods.

The benchmarking research shows that, on all assessment measures, the suggested STFO-FCMC method performs noticeably better than current state-of-the-art techniques like LGB and stacking. In particular, the accuracy, precision, recall, and $F1$-score of STFO-FCMC were 97.84, 96.02, and 95.94%, respectively. The LGB model performed worse, with an accuracy of 91.24%, a precision of 88.06%, a recall of 87.46%, and an $F1$-score of 87.76%, whereas the stacking model recorded an accuracy of 94.59%, a precision of 92.18%, a recall of 90.17%, and an $F1$-score of 91.16%. These findings demonstrate STFO-FCMC's strong potential for use in tourist analytics by confirming its supremacy in capturing intricate patterns of customer behavior.

4.3 Cluster analysis on tourism consumer behavior

Clustering analysis in tourism consumer behavior is a technique for categorizing a population of tourists into discrete group clusters based on similarities in their travel choices and activities. Figure 5 reveals the outcome of cluster analysis.

Cluster 1 prefers adventure (40%) and relaxation (20%), cluster 2 values both adventure and relaxation (30%), cluster 3 prioritizes luxury (50%), cluster 4 enjoys city travel (50%), and cluster 5 focuses on family (60%).
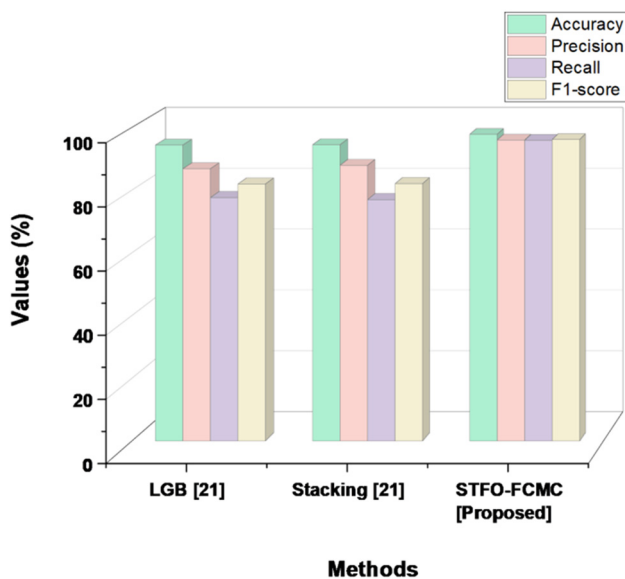

Figure 4: Result of performance metric.

Table 3: Outcomes of comparative analysis

Methods	Accuracy (%)	Precision (%)	Recall (%)	F1-score (%)
LGB [21]	94.41	86.84	77.63	81.97
Stacking [21]	94.50	87.94	76.97	82.09
STFO-FCMC [proposed]	97.84	96.02	95.94	96.19

4.4 Business performance

Business performance is the extent to which a corporation achieves its objectives, as evaluated by customer happiness, revenue growth, customized service adoption, marketing campaign performance, and overall operational efficiency. Customer satisfaction rate is computed by averaging user review scores on a scale of 1–5 and scaling to a percentage. Personalized service adoption is determined by the percentage of users who opted into customized package recommendations out of the total user base. The conversion rate is computed as the ratio of completed bookings to total website or app visits. Customer retention rate is measured as the percentage of users making repeat bookings within a set timeframe, identified via unique user IDs. Revenue growth from custom packages is estimated by comparing total revenue from personalized offers before and after the implementation of STFO-FCMC. Marketing campaign effectiveness is assessed by measuring the increase in user engagement directly linked to specific campaigns over time. Figure 6 illustrates the result of business performance.

The customer satisfaction rate is 89.3%, personalized service adoption is 85%, the conversion rate is 50.6%, customer retention rate is 81.7%, revenue growth from custom packages is 27.9%, and marketing campaign effectiveness is 70.5%. Customer satisfaction shows positive outcomes, with marketing campaign effectiveness contributing significantly to overall performance.

4.5 Consumer satisfaction

In tourism, fuzzy clustering analysis refers to the amount of satisfaction visitors have with personalized services,
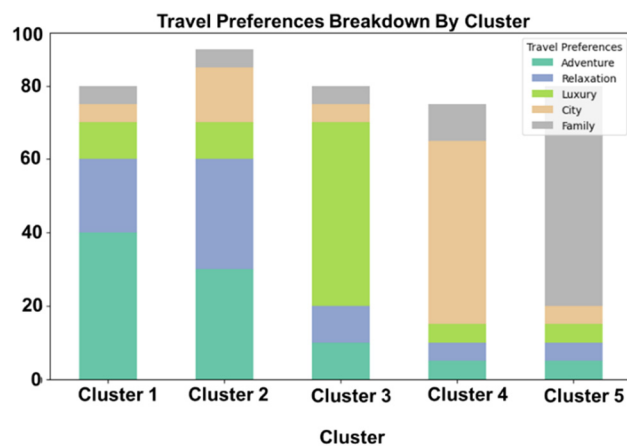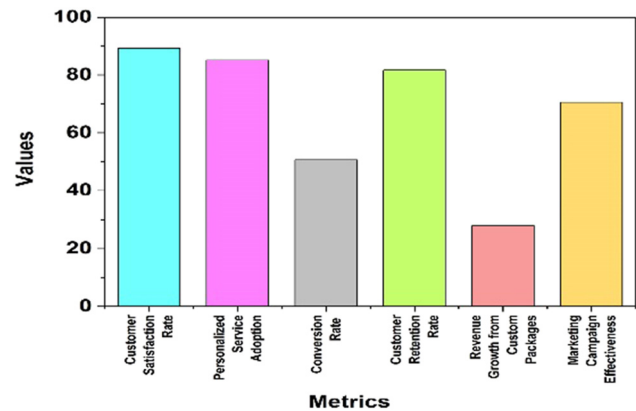
Figure 6: Result of business performance.

bespoke trip packages, and marketing methods that fit their interests, eventually improving their entire travel experience and loyalty. Figure 7 shows the outcome of consumer satisfaction.

4.6 Efficiency in consumer behavior analysis

It refers to an analytical model's or algorithm's capacity to effectively analyze and interpret consumer data, with a focus on both performance and resource utilization. High efficiency means that the model provides relevant insights quickly and consistently while using minimal computing resources. Table 4 shows the efficiency of consumer behavior analysis.

The clustering accuracy of 91.8%, Davies Bouldin Index (DBI) of 1.38, and silhouette coefficient of 0.74 indicate well-separated distinct clusters. With 9.1 s execution time,
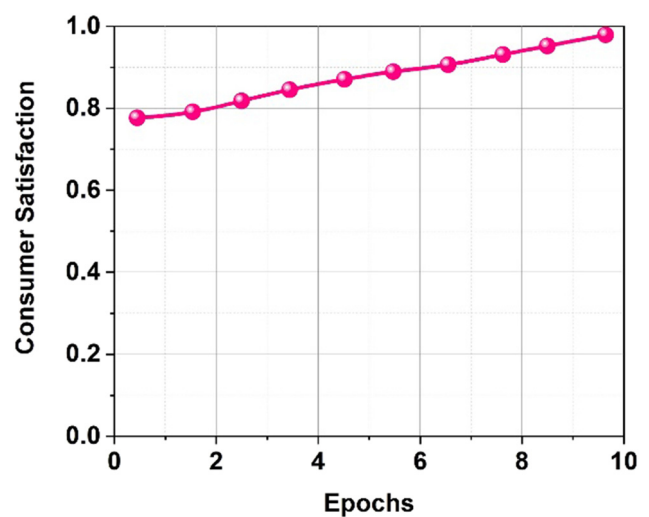
Figure 5: Outcome of cluster analysis.

Figure 7: Outcome of consumer satisfaction.

Table 4: Efficiency in consumer behavior analysis

Metric	STFO-FCMC
Clustering accuracy	91.8
DBI	1.38
Silhouette coefficient	0.74
Execution time (s)	9.1
Membership value consistency	0.84
Computational complexity (O)	$O(n \log n)$
Convergence iterations	18
Memory usage (MB)	270

Table 2: Experimental setup

Component	Details
Hardware	Intel Core i7,16 GB RAM,1 TB SSD
Operating system	Windows 10/11, Linux (Ubuntu, CentOS)
Software	Python, R
Clustering algorithm	FCM
Libraries/packages	sci-kit-learn, fuzzy-c-means, pandas, NumPy
Visualization tools	Seaborn, plotly
IDE/environment	Jupyter Notebook

membership value consistency of 0.84, O ($n \log n$) complexity, then 18 iterations and 270 MB memory usage, the model balances performance and efficiency. Figure 8 displays the outcome of the membership matrix.

The membership matrix fuzzy c means clustering for tourism behavior shows the membership degree of each data point to the three clusters. Cluster 1, budget travelers; cluster 2, luxury travelers; and cluster 3, adventure seekers. The color bar shows the membership degree, demonstrating how strongly each point belongs to each cluster.

5 Discussion

The goal is to use fuzzy cluster analysis to detect tourist activity patterns and classify unique groups while offering insights to enhance tourism marketing strategies and services. The SVM [8] approach has the drawback of a simulated dataset, which did not adequately replicate real-world settings. DL [13] model's performance is influenced

by biased review data, domain-specific constraints, and challenges in handling multilingual or nuanced sentiments. It is limited to existing literature and lacks empirical validation through real-world case studies or experimental data [15]. The NN approach lacks real-world implementation validation and does not account for rapidly changing consumer behaviors [10]. The proposed method was compared with traditional methods like LGB [21] and stacking [21] for consumer behavior analysis. LGB [21] has interpretability issues, problems with extremely unbalanced data, and overfit small datasets. Stacking [21] was difficult to install and fine-tune, computationally costly, and prone to overfitting without adequate validation. Because of their processing cost, both needed to be handled carefully and were less appropriate for real-time applications. STFO-FCMC overcame the limitations of prior approaches by offering greater scalability, increasing tourist customer behavior through company success and satisfaction, and rapidly managing datasets to solve this issue.

6 Conclusion

The efficacy of fuzzy clustering was investigated in this research, specifically the STFO-FCMC method, to identify diverse consumer behavior patterns in the tourism industry. Fuzzy clustering was combined with consumer data, including travel preferences and demographics, to uncover important behavioral trends that were used by travel agencies to create customized services and focused marketing campaigns that will improve customer satisfaction and overall business performance. The STFO-FCMC technique demonstrated good clustering quality using labeled subsets and internal validation metrics, producing high clustering alignment with known behavior patterns (accuracy of 97.84%, precision of 96.02%, recall of 95.94%, and F1-score of 96.19%). These metrics were calculated using labeled data for benchmarking purposes, but the
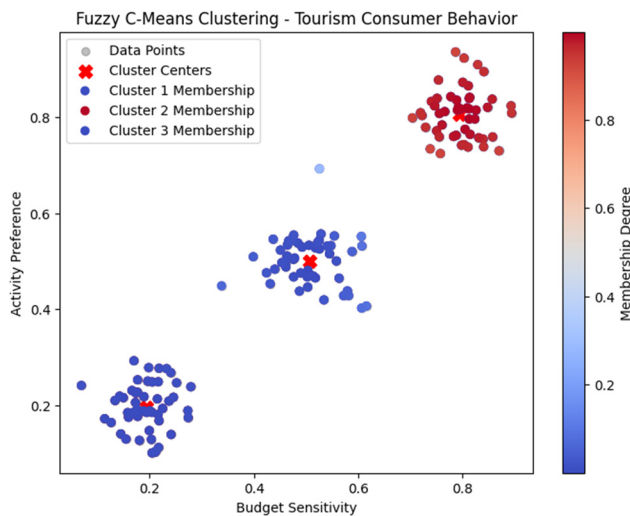
Figure 8: Outcome of the membership matrix.

research has limitations, such as a reliance on self-reported data that was susceptible to bias and a dataset that was limited in size and diversity, which may limit the generalizability of findings. Future work will focus on incorporating real-time behavioral data and exploring alternative advanced clustering techniques to further improve consumer segmentation and decision-making support in diverse tourism markets.

References

[1] M. N. Cunha, M. Pereira, A. Cardoso, J. Figueiredo, and I. Oliveira, "Redefining consumer engagement: The impact of AI and machine learning on marketing strategies in tourism and hospitality," *Geo J. Tour. Geosites*, vol. 53, no. 2, pp. 514–521, 2024, doi: 10.30892/gtg. 53214-1226.

[2] M. J. Kim, C. M. Hall, O. Kwon, and K. Sohn, "Space tourism: Value-attitude-behavior theory, artificial intelligence, and sustainability," *J. Retail. Consum. Serv.*, vol. 77, p. 103654, 2024, doi: 10.1016/j. jretconser.2023.103654.

[3] S. S. Kim, W. Shin, and H. W. Kim, "Unravelling long-stay tourist experiences and satisfaction: text mining and deep learning approaches," *Curr. Issues Tour.*, vol. 28, no. 3, pp. 492–510, 2024, doi: 10.1080/13683500.2024.2327840.

[4] R. Sann, P. C. Lai, S. Y. Liaw, and C. T. Chen, "The nature of electronic complaints about dark tourism destinations: A machine learning approach," *J. Heritage Tour.*, vol. 20, no. 1, pp. 107–129, 2025, doi: 10.1080/1743873X.2024.2399208.

[5] D. Shrestha, T. Wenan, D. Shrestha, N. Rajkarnikar, and S. R. Jeong, "Personalized Tourist recommender system: a data-driven and machine-learning approach," *Computation*, vol. 12, no. 3, p. 59, 2024, doi: 10.3390/computation12030059.

[6] S. Blanco-Moreno, A. M. González-Fernández, and P. A. Muñoz-Gallego, "Big data in tourism marketing: past research and future opportunities," *Span. J. Mark.*, vol. 28, no. 3, pp. 266–286, 2024, doi: 10.1108/SJME-06-2022-0134.

[7] B. S. Al-Romeedy and T. Hashem, From insight to advantage: harnessing the potential of marketing intelligence systems in tourism, *In Marketing and Big Data Analytics in Tourism and Events*, IGI Global, 2024, pp. 80–98. doi: 10.4018/979-8-3693-3310-5.ch005.

[8] H. Ma, "Development of a smart tourism service system based on the Internet of Things and machine learning," *J. Supercomputing*, vol. 80, no. 5, pp. 6725–6745, 2024, doi: 10.1007/s11227-023-05719-w.

[9] H. S. Saragih, M. R. U. Saputra, and M. H. Dewantara, "Exploring topics and trends in service robots, artificial intelligence, and realities in tourism: a text-mining approach," *Emerging Technologies in Business: Innovation Strategies for Competitive Advantage*, Singapore: Springer; pp. 239–259, 2024, doi: 10.1007/978-981-97-2211-2_11.

[10] H. Liu, "Big data precision marketing and consumer behavior analysis based on fuzzy clustering and PCA model," *J. Intell. Fuzzy Syst.*, vol. 40, no. 4, pp. 6529–6539, 2021, doi: 10.3233/JIFS-189491.

[11] X. Zhang, M. Cheng, and D. C. Wu, "Daily tourism demand forecasting and tourists' search behavior analysis: a deep learning approach," *Int. J. Mach. Learn. Cybern.*, pp. 1–14, 2024, doi: 10.1007/s13042-024-02157-9.

[12] S. Yang, Q. Li, D. Jang, and, J. Kim, "Deep learning mechanism and big data in hospitality and tourism: Developing personalized restaurant recommendation model to customer decision-making," *Int. J. Hosp. Manag.*, vol. 121, p. 103803, 2024, doi: 10.1016/j.ijhm. 2024.103803.

[13] L. Meng, "The convolutional neural network text classification algorithm in the information management of smart tourism based on Internet of Things," *IEEE Access*, vol. 12, pp. 3570–3580, 2024, doi: 10.1109/ACCESS.2024.3349386.

[14] Y. A. Singgalen, "Culture and heritage tourism sentiment classification through cross-industry standard process for data mining," *Int. J. Basic. Appl. Sci.*, vol. 12, no. 3, pp. 110–120, 2023, doi: 10.35335/ijobas.v12i3.299.

[15] E. Cherenkov, V. Benga, M. Lee, N. Nandwani, K. Raguin, M. C. Sueur, et al., "From machine learning algorithms to superior customer experience: business implications of machine learning-driven data analytics in the hospitality industry," *J. Smart Tour.*, vol. 4, no. 2, pp. 5–14, 2024, doi: 10.52255/smarttourism. 2024.4.2.2.

[16] E. Bartl, M. Weigert, A. Bauer, J. Schmude, M. Karl, and H. Küchenhoff, "Understanding travel behavior patterns and their dynamics: Applying fuzzy clustering and age-period-cohort analysis on long-term data of German travelers," *Eur. J. Tour. Res.*, vol. 39, pp. 3914–3914, 2025, doi: 10.54055/ejtr.v39i.3862.

[17] M. A. Imran and K. Hyun, "A novel pattern recognition technique to characterize multi-day shopping and entertainment trip activities," *Travel. Behav. Soc.*, vol. 40, p. 101035, 2025, doi: 10.1016/j.tbs.2025. 101035.

[18] K. Chrysafiadi, A. Kontogianni, M. Virvou, and E. Alepis, "Enhancing user experience in smart tourism via fuzzy logic-based personalization," *Mathematics*, vol. 13, no. 5, p. 846, 2025, doi: 10.3390/math13050846.

[19] M. Sağbaş and S. Aydogan, "Unveiling the nuances: how fuzzy set analysis illuminates passenger preferences for ai and human agents in airline customer service," *Tour. Hosp.*, vol. 6, no. 1, p. 43, 2025, doi: 10.3390/tourhosp6010043.

[20] https://www.kaggle.com/datasets/zoya77/tourism-consumer-behavior-insights-dataset.

[21] S. X. Chen, X. K. Wang, H. Y. Zhang, J. Q. Wang, and J. J. Peng, "Customer purchase forecasting for online tourism: A data-driven method with multiplex behavior data," *Tour. Manag.*, vol. 87, p. 104357, 2021, doi: 10.1016/j.tourman.2021.104357.